Open Source Software (Digital Libraries): With Special Reference to Koha and Dspace

Dr. P. Venkateshwar Rao
Assistant Professor (Assistant Librarian), LIS, University College of Science, Osmania University, Andhra Pradesh, India
B. Kumar
Library Assistant, University College of Engineering, Osmania University, Andhra Pradesh, India

Abstract:
Open source software in libraries, with description of a variety of software packages and successful library projects. There are many shared principles between Open source software and Librarianship, especially the free and equal access to information. Because of the freedom we gain with the use of Open source software is it possible to have greater control over how libraries operate. Anybody who worked with computers on a daily basis can contribute to open source software because things like information architecture, usability testing, documentation, and staffing are key skills required for successful projects, and these skills are inherent in the people who use computers as a primary tool in the work.
Koha is a full featured Integrated Library System (ILS) specially meant for automation of library services and activities. DSpace is an Open Source Software package that provides the tools for management of digital assets, and is commonly used as the basis for an institutional repository.
Now the library and information sectors have become totally users oriented so the concept of 'customer's satisfaction' also applies here and Open source software is essential in libraries to fulfill the user's needs in appropriate manner. This paper highlights how the libraries are highly influenced by the digital resources especially the internet usage by the user community to retrieve information for various. This also highlights various features of the Open source software for Library Automation, Open source software for Digital Library and Institutional Repositories and Open source software for content management in libraries.

Keywords: Open Source Software, Open Source Digital Library System, Open Source Software for Library Automation – KOHA, Digital Libraries, Institutional Repositories, Digital Library Software – Dspace, Library Software.

1. Introduction

Librarians who are looking for alternatives to restrictive software licenses and expensive software, which does not fully meet their needs, continue to incorporate open source software into their institutions. In the past years there have been a number of developments in library related open source software.

Simply stated, open source software to supply with that source code, the underlying programming which is used to create any software package. In the case of proprietary software, the end user cannot legally view or change the source code By contrast, Open source software users are encouraged to look at the source code and offer improvement where possible, using a process which is similar to peer review.

Open source software is both a philosophy and a process. As a philosophy it describes the intended use of software and methods for its distribution. Depending on your perspective, the concept of Open source software is a relatively new idea being only four or five years old. On the other hand, the GNU Software Project – a Project advocating the distribution of "free" software – has been operational since the mid 80's. Consequently, the ideas behind Open source software have been around longer than y7ou may think. It begins when a man named Richard Stallman worked for MIT in an environment where software was shared. In the mid 80's Stallmen resigned from MIT to begin developing the GNU – a software project intended to create an operating system much like UNIX. (GNU is pronounced "gnu-NEW" and is a recursive acronym for GNU's Not UNIX). His desire was to create "free" software, but the term "free" should be equated with freedom, and as such people who use "free" software should be:

- Free to run software for any purpose.
- Free to modify the software to suit their needs.
- Free to redistribute of the software gratis or for a fee.
- Free to distribute modified versions of the software.

The following are the Open source software's for the integrated library system:
Open source software for Library Automation.

- Koha
- Evergreen
- Open Biblio
- PhpMyLibrary

Open source software for Digital Library and Institutional Repositories.

- DSpace
- E-prints
- GSDL

Open source software for content management in libraries.

- Drupal
- Joomla
- Plone

2. Why Is Open Source Important?

The open source is important because it underpins both Web 2.0 and the smart mobile devices accessing and interacting with the cloud at work and socially. It is supporting fantastic amounts of innovation, innovation that is changing people's lives. I believe that the future includes, intuitive to use always on low power smart devices.

2.1. Need of Open Source Software

Price of propriety library management software's is very high and financial weak libraries cannot invest large amount for library automation. Extra payment is required for software updating and maintenance in future. Open source Library Management software's consist the entire essential functional module which were available with proprietary software's. Features are as fallows.

1. Open License (General Public License (GPL))
2. Open Source ILS (Integrated Library Software) are operating systems independent. They are compatible with Linux, Windows and McOS.
3. Web interface
4. MARC support

2.2. Benefits of Using Open Source Software

The outline flow of this section is as follows: Organisational benefits from the use of Open Source Software - Reliability, Stability, Auditability, Cost, Flexibility and Freedom, Support and Accountability

2.3. Background and Purpose

In the earliest stages of the transition from a CD-Rom-based collection to the WWW site, it was clear that the nature and scope of the Perseus resource demanded a flexible, extensible, and powerful data management system. Written mostly in Perl, the production version of the on-line Perseus text management system evolved and grew over eight years, becoming a uniquely powerful platform, capable of ingesting heterogeneous source materials and performing a range of automatic services.

2.4. Open – Source Services

The Perseus Hopper is an open-source project providing a suite of services for interacting with textual collections. While as a whole it provides an integrated reading environment, its individual services are designed to be modular and can be grouped into three different classes.

2.4.1. Linguistic Support

The Hopper itself is language independent, but the code includes native support for Greek, Latin and Arabic. Given a source text in any one of those three languages (either a text bundled with the code release or a TEI-compliant XML text of the user's own), it provides services for automatic lemmatization (linking inflected word forms to the dictionary entries from which they're derived) and morphological analysis (identifying, for instance, that the Latin word *amor* is a singular masculine nominative noun). At a broader level, it also enables corpus research by automatically generating word and lemma frequency information for the entire collection of texts supplied to it.

2.4.2. Contextualized Reading

Users can take advantage of this contextualization with the Greco-Roman and Arabic texts provided, or specify themselves the higher-level relationship between their own texts (e.g., that document X *is a translation of* document Y) in order to create a reading environment where passages in a source text are accompanied by secondary resources such as translations and commentaries.

2.4.3. Searching

Users can not only read passages from texts, but use a suite of search tools to find what they are looking for, in any of the languages the Hopper supports. These search tools include word and phrase searches, in individual texts or collections. These searches include the option to search all possible inflections of a word, making them extremely powerful for morphologically rich languages like Greek, Latin and Arabic (e.g., a lemmatized search for the root form *sum* would also find documents containing the inflected forms *est* and *sunt*).

2.5. Extensibility
The code base itself invites two varieties of extensibility. On one hand, while the code is bundled with a collection of Greco-Roman and Arabic texts around which it has grown, users are able to include their own TEI-compliant XML texts as part of the reading environment and enable the same services for those texts as those that are available online for Perseus' open-source editions.   On the other hand, the Java code itself is also designed with modularity and extensibility in mind.   An example of this is the variety of classes (all ultimately inherited from Corpus Processor) to cycle through an entire collection of texts and performs some operation on each one.

2.6. Examples of Open Source Software
- Linux a computer operating system and kernel originating as a UNIX system and existing in many versions.
- GNOME a desktop development environment providing tools for developing desktop applications and available for Linux and other UNIX systems.
- KDE another desktop development environment.
- Apache a web server program used in most web server computers and available for Windows as well as Linux and other UNIX systems.
- Firefox a web client program (for viewing and hearing web pages) available for Windows and Mac as well as Linux and other UNIX systems.
- Dovecot an email server program (for sending and receiving email) supporting POP3 and IMAP and available for Linux and other UNIX systems.
- Postfix (see http://www.postfix.org): an email server program (for sending and receiving email) supporting SMTP and available for Linux and other UNIX systems.
- SquirrelMail an email application (for reading and writing email through a web client program) supporting IMAP and SMTP and available for Windows and Mac as well as Linux and other UNIX systems.
- Thunderbird (see Error! Hyperlink reference not valid.): an email client program (for reading and writing email) supporting POP3, IMAP and SMTP and available for Windows and Mac as well as Linux and other UNIX systems.
- OpenOffice (see Error! Hyperlink reference not valid.): an office application set (for word processing, spreadsheet manipulation, picture drawing and data base access) compatible with all other major office application sets such as Office and available for Windows and Mac as well as Linux and other UNIX systems.
- KOffice: (see Error! Hyperlink reference not valid.): another office application set.
- Asterisk: (see Error! Hyperlink reference not valid.): a Private Branch eXchange (PBX) providing features such as interactive voice response, conference bridging and call queuing (as well as analogue, digital and VoIP telephony), supporting H.323, SIP and IAX2 (and other VoIP protocols), and available for Mac as well as Linux and other UNIX systems.
- FreeSWITCH (see Error! Hyperlink reference not valid.): a VoIP development environment providing tools for developing applications like phones and switches, supporting H.323, SIP and IAX2 (and other VoIP protocols), and available for Windows and Mac as well as Linux and other UNIX systems.
- GnuGK (see Error! Hyperlink reference not valid.): a VoIP gatekeeper (for some H.323 deployments) available for Windows and Mac as well as Linux and other UNIX systems.

2.7. Advantages and Disadvantages of Open Source Software?
Open source is a fairly new concept that has gained huge popularity in the field of IT in recent years. This is mainly because open-source software is free to use – its greatest advantage. As it is developed by a non-profit community, it has some disadvantages as well.

2.7.1. Advantages
i. Open-source software is free to use, distribute, and modify. It has lower costs, and in most cases this is only a fraction of the cost of their proprietary counterparts.
ii. Open-source software is more secured as the code is accessible to everyone. Anyone can fix bugs as they are found, and users do not have to wait for the next release.
iii. The fact that is continuously analyzed by a large community produces secure and stable code.
iv. Open source is not dependent on the company or author that originally created it. Even if the company fails, the code continues to exist and be developed by its users. Also, it uses open standards accessible to everyone; thus, it does not have the problem of incompatible formats that exist in proprietary software.

    v.   Lastly, the companies using open-source software do not have to think about complex licensing models and do not need anti-piracy measures like product activation or serial number.

<u>2.7.2. Disadvantages</u>

- The main disadvantage of open-source software is not being straightforward to use. Open-source operating systems like Linux cannot be learned in a day. They require effort and possibly training from your side before you are able to master them. You may need to hire a trained person to make things easier, but this will incur additional costs.
- There is a shortage of applications that run both on open source and proprietary software; therefore, switching to an open-source platform involves a compatibility analysis of all the other software used that run on proprietary platforms. In addition, there are many ongoing parallel developments on open source software. This creates confusion on what functionalities are present in which versions.
- Lastly, many of the latest hardware are incompatible to the open-source platform; so you have to rely on third-party drivers.
- The decision of adopting open-source software should not be taken just on the basis of the low-cost involved. It entails a detailed analysis and understanding of the requirements before switching to open source to get full benefits of it.

3. Digital Libraries

Libraries have taken up digitization process to maximize the use and accessibility of the resources by facilitating better access to educational, cultural, & scientific materials. Advances in information technology and growth of data in digital formats have also lead to the creation of various methods to assist the users in locating desired data. Most significant shift in building digital resources is greater interoperability among information systems across the country and world over.   Building the digital collections and infrastructure required to access them is a challenge that every library has to face Libraries have expanded with the developments in information technology and creation of digital information resources.   Information Technology tools provide opportunities for the libraries to store, process, and transfer and exchange information.

Digital libraries are built around Internet and web technologies with electronic databases/e-documents and e-content as their building blocks.   Increased usage of software, hardware and networking technologies, WWW and also highly evolved browsers have paved way for the creation of digital libraries at global level.

3.1. Open Source Software for Digital Library Management

The Digital Library Management software's (DLMS) present an easy to use, customizable architecture to create online digital libraries. With these institutions/organizations can disseminate their research work, manuscripts, or any other digital media for preservations and world over dissemination of digital items. The software's discussed above present different services and architectures. It is difficult to propose one specific DLMS system as the most suitable for all cases. The study can be used as a reference guide by any organization or institute to decide which one will be I deal for creating and showcasing their digital collection. The choice usually depends on type/format of material, distribution of material, software platform and time frame etc for setting up a Digital Library

<u>3.1.1. DSPACE</u>

DSpace is the most popular among the digital library solutions available in the open source main. Eprints is also widely used. Educational institutions dominate in the use of these packages. Though many institutions have implemented digital libraries, only about half of them are available online. Open access of knowledge is possible only if these repositories are made online. India is benefiting well from the open source movement. It is difficult to propose one specific DL system as the most suitable for all cases. Each system has its advantages and drawbacks, as stated in the above comparison, categorized by basic DL system characteristics and features. That comparison can only be used as a guideline by an organization in order to decide if one of these DL systems is suitable to host its digital collections. Usually the needs for each organization vary depending on the number of collections, the types of objects, the nature of the material, the frequency of update, the distribution of content and the time limits for the development of a DL. In the next paragraphs, guidelines for the selection of a DL system are provided depending on different organization needs.

1. Consider a case where an institution or university needs a digital repository for research papers and dissertations produced by students and stuff. In that case, the most appropriate DL system is DSpace, since it by default represents communities (e.g. university departments) and collections (e.g. papers and dissertations), while workflow management supported is important for item submission by individuals.
2. Consider a case where an organization needs one digital collection to publish its digital content in a simple form, in strict time limits. In addition, the organization prefers to integrate the web interfaces of the DL with a portal like website. In that case the most appropriate DL systems are Keystone or EPrints, since they separate the concerns of presentation and storage, are not bind to specific metadata standards and provide simple web interfaces for the submission and presentation of documents and metadata.
3. Consider a case where an organization is responsible to digitize collections from libraries, archives and museums and host them in a single DL system. The organization has human resources and the amount of time in order to customize the DL system and develop extra modules. The highest priority needs are the support of preservation issues, the use of multiple metadata standards and the different formats of digital content. In that case the most suitable DL system is Fedora, since it

provides a very customizable modular architecture. Although it does not provide easy to use web interfaces or built-in functionality, it is the best choice for the case where many collections and different material must be hosted.

4. Consider a case where an organization wants to electronically publish books in an easy to use customizable DL system. In that case the most appropriate DL system is Greenstone, since it is easy to represent books in a hierarchical manner, using table of contents, while the full text of chapters can be searchable.

3.2. Open Source Software for Library Automation

3.2.1. Koha
Koha was developed in 1999 by a small team of programmers working for a consulting company in New Zealand to address the needs of a small library branch on the island because their vendor supplied ILS was outmoded and not Y2K compliant. The library was also faced with rising phone line costs as their older system depended on modems to maintain its "sneaker" network. The initial goals ware to duplicate functionally and maintain ease of use with TCP/IP connectivity and web base GNU staff terminals

3.2.2. Software Evaluation with Checklists
To effectively evaluate the chosen LMS software, use of checklists is necessary. Checklists have been used widely to verify the correctness of software documentation and software code. Gilb and Graham defined as "a specialized set of questions designed to help checkers find more defects".
The Acquisition, Catalogue, Circulation, Serial Control, Web OPAC and Reports modules are available in Koha and NewGenLib software. Acquisition module is not available in Openbiblio and phpMyLibrary software. Whereas Serial control module is not available in OpenBiblio, and Reports module is not available in PhpMyLibrary software.

4. Conclusion
The Library & Information Science (LIS) professionals should keep eyes on development in order to choose appropriate technology depending upon Institution's needs. Since, numbers of libraries worldwide are using OSS for managing their library systems more economically and effectively. Librarians and programmers should worked together in order to implement open source integrated library systems and at the same time, library professional are also required to acquire new skills for developing and managing the library by using open source LMS. For taking benefit from OSS additional technology, education, and training of the professionals is essentially required.
The 21st century has come out with convergence of wireless technology, fiber optics, software application and new generation Internet switches, IP version that will permit anything with electricity to have a web address and run off the internet. Open source software are best alternative to the costly library software available in market. Open source software have the professional and working standard. They have all the modules which are required for library automation and digital library. Before going for the costly software one must try an open source software.

5. References
i. C. Lagoze and H. Van de Sompel. The Open Archives Initiative: Building a low-barrier interoperability framework. In Proceedings of the Joint Conference on Digital Libraries (JCDL '01), 2001.
ii. DSpace Federation. Available at http://www.dspace.org/
iii. R. Kahn and R. Wilensky. A Framework for Distributed Digital Object Services. Corporation of National Research Initiative - Reston USA, 1995. Available at http://www.cnri.reston.va.us/k-w.html
iv. DCMI Metadata Terms. Dublin Core Metadata Initiative. Available at http://www.dublincore.org/documents/dcmi-terms/
v. METS: An Overview & Tutorial. Library of Congress. Available at http://www.loc.gov/standards/mets/METSOverview.v2.htm
vi. http://www.oss4lib.org/node/8
vii. http://www.perseus.tufts.edu/hopper/
viii. http://www.hrsa.gov/healthit/toolbox/HealthITAdoptiontoolbox/OpenSource/softwareadvantage.html
ix. http://arxiv.org/abs/1212.4935
x. http://commons.gc.cuny.edu/wiki/index.php/Open_Source_-_Digital_Libraries
xi. http://www.oss4lib.org/node/8
xii. http://eu.conecta.it/paper/Advantages_open_source_soft.html
xiii. http://www.cloudtweaks.com/2012/08/advantages-and-disadvantages-of-open-source/